ISSN 2029-6991 / eISSN 2029-7009

VILNIUS TECH
Vilnius Gediminas Technical University

GEODESY and CARTOGRAPHY

2024
Volume 50
Issue 4
Pages 162–178
https://doi.org/10.3846/gac.2024.20106

UDC 528.714

TEMPORAL ANALYSIS OF MULTI-SPECTRAL INSTRUMENT LEVEL AND SURFACE REFLECTANCE DATA SETS FOR SEASONAL VARIATION IN LAND COVER DYNAMICS BY USING GOOGLE EARTH ENGINE

Anubhava SRIVASTAVA✉

School of Engineering and Technology, Department of Computer Science and Engineering, Sharda University, Greater Noida, Uttar Pradesh, India

Article History:
- received 08 August 2023
- accepted 03 December 2024

Abstract. By rapid growth in programming tools, accessibility to end consumer computing power, and the availability of free satellite data, the data science and remote sensing fields have begun to converge in recent years. Before this major processing time is wasted in collection of data. Google Earth Engine easily overcomes above problem; it contains data from different satellites and has power of processing and computation also. Well known data provider satellites are present in the library of GEE and users can easily process and track real time data from these satellites over GEE. "Sentinel", a mission of the European Space Agency and "Landsat", an American Earth observation satellite have been used in a variety of remote sensing applications. GEE makes these data sets available to the general public. These datasets are utilised for computing and analysis purposes. The objective of this study is to find change in study area by using above discussed two satellite data, over each season of year on different category of classification (Random Forest, CART, GTB and SVM). This work focuses on improving the classification accuracy of different classification algorithm by reviewing training samples and analyzing post-classification with image differencing in the algebraic technique. Because Landsat data have a medium spatial resolution, therefore point-wise computation was used. Lastly, we also detect which data sets are working better on an appropriate machine learning algorithm, so after final calculation we estimate accuracy of each algorithm by using confusion matrix and kappa.

Keywords: GEE, remote sensing, classification, Landsat, Sentinel, satellite data.

✉Corresponding author. E-mail: ***pgi18001@rgipt.ac.in***

1. Introduction

Large-scale land cover information is pivotal for resource management, policy formation, and human activities. Global environmental and sustainability studies emphasize land cover changes as a critical factor, influenced by various socioeconomic, environmental, political, technological, and cultural conditions. Statistical-based modeling is commonly used to evaluate temporal changes in land cover but often lacks spatial scale changes. Remotely sensed data have become invaluable for understanding land cover changes due to their ability to encompass wide temporal and spatial scales. However, classifying land cover from satellite imagery remains challenging. Typically, this involves creating a library of data from various sources such as USGS, Copernicus, NASA, ISRO, etc., importing these data into GIS tools like QGIS or ARCGIS, merging datasets for improved evaluation, and classifying them while validating accuracy against ground truth data. Google Earth Engine offers a solution to these challenges by providing a cloud-based platform with a collection of satellite data from different sensors. This paper aims to evaluate Google Earth Engine's capacity as a web-based remote sensing platform for spatial and temporal aggregations of satellite images. Focusing on Dehradun as a study area, we assess land cover changes over time, considering the computational complexities and challenges within GEE. Many researcher have applied different mechanisms (Goldblatt et al., 2017; Qiao et al., 2019; A. Srivastava et al., 2022; Viana et al., 2019) for computing change in area by using Land Use Land Cover (C. Liu et al., 2019; Y. Liu et al., 2020) classification technique. The remainder of the paper is structured as follows: Section 2 states previous studies using the Google Earth Engine over world wide and study areas. Section 3 is divided into three subsections. The first subsection delineates the study area, while the second subsection discusses the various datasets utilized for analysis purposes, and the final subsection delves into the accuracy assessment of

performance analysis. This subsection involves calculating accuracy using various well-known algorithms. The results are detailed in Section 4. Lastly, Section 5 offers concluding remarks.

2. Related work

Numerous studies have investigated changes in land cover using remote sensing data and various analytical techniques. One notable approach is the use of multispectral satellite imagery to monitor and analyze land cover changes over time. For instance, Serwa and Elbialy (2021), Srivastava and Biswas (2023) utilized Landsat imagery to assess land cover changes in a tropical forest landscape, highlighting the effectiveness of remote sensing for monitoring deforestation and forest degradation (Srivastava et al., 2023c). Similarly, Jin et al. (2022), Serwa (2012), Srivastava et al. (2023b) employed Sentinel-2 imagery to detect urban expansion and land cover changes in rapidly developing regions, emphasizing the importance of high-resolution data for urban planning and environmental management. Furthermore, researchers have explored the application of surface reflectance data for land cover analysis. Gupta (2015) investigated changes in land cover patterns using surface reflectance data from Landsat 8, revealing significant alterations in forest cover due to anthropogenic activities. Additionally, Serwa et al. (2010) utilized Google Earth Engine (GEE) (Campos-Taberner et al., 2018; Gorelick et al., 2017) to analyze land cover changes in agricultural landscapes, demonstrating the utility of GEE for large-scale land cover assessments. Similarly, Serwa and Elbialy (2021) compared the performance of the two sensors for monitoring urban expansion, noting the superior spectral resolution of Sentinel-2 for discriminating urban land cover classes and capturing urban dynamics over time. Furthermore, researchers have investigated the influence of seasonal variations on the performance of Sentinel-2 and Landsat 8 data for land cover analysis. Benediktsson et al. (2007) evaluated the impact of seasonal changes on vegetation indices derived from both sensors, highlighting the importance of considering seasonal variability in satellite data interpretation. Additionally, Thenkabail et al. (2021), Serwa and Elbialy (2021), Srivastava et al. (2023a) examined the seasonal dynamics of water bodies using Sentinel-2 and Landsat 8 imagery, demonstrating the potential of both sensors for monitoring changes in water resources throughout the year. Later Srivastava and Sharma (2024) uses Google Earth Engine and different datasets for observing environmental changes also.

While these studies have provided valuable insights into the comparative analysis of Sentinel-2 and Landsat 8 data, there remains a need for comprehensive assessments that consider multiple factors, including spectral characteristics, spatial resolution, and temporal coverage, across different seasons. In this study, we aim to address this gap by performing a detailed comparison of Sentinel-2 and Landsat 8 datasets over different seasons of the year. By evaluating the performance of both sensors under varying environmental conditions, we seek to provide valuable guidance for researchers and practitioners in selecting the most appropriate satellite data for their specific applications.

3. Data and methods

3.1. Study area

The study area, nestled in the Doon Valley of the Himalayan foothills, lies between the Song River, a tributary of the Ganga, to the east, and the Asan River, a tributary of the Yamuna, to the west. This region serves as a gate-

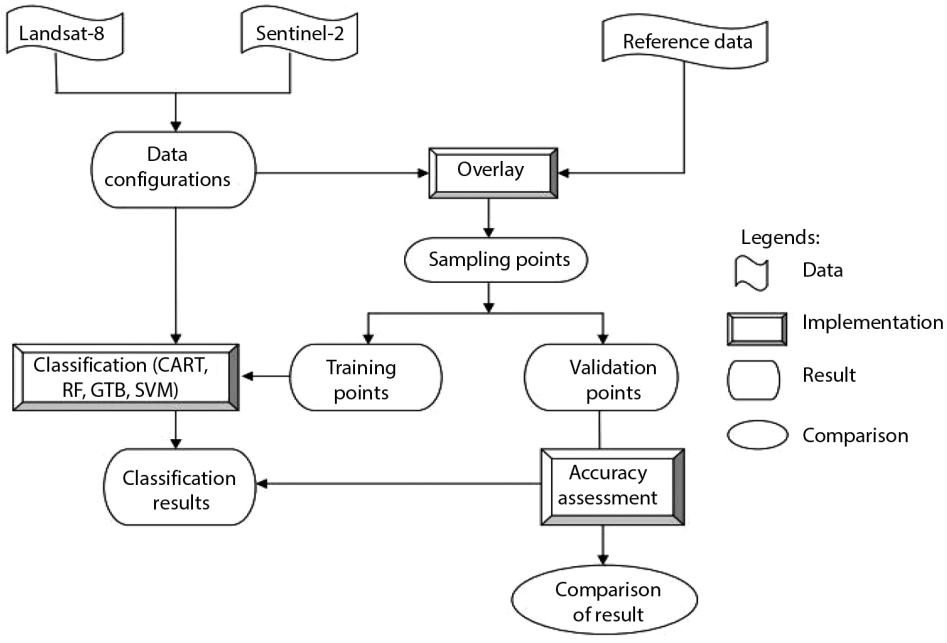


Figure 1. Working culture adopted in computation

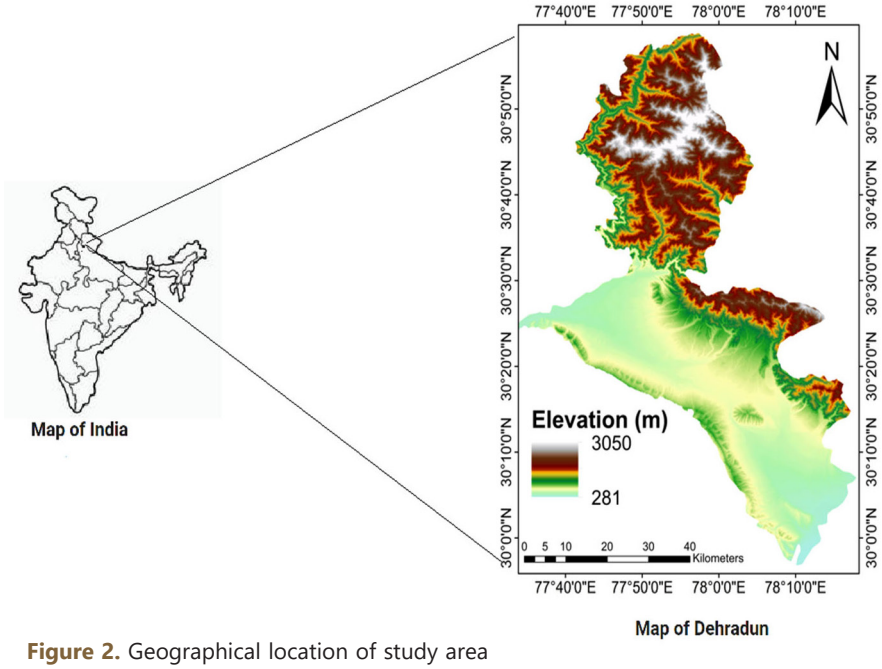


Figure 2. Geographical location of study area

way to the surrounding areas, boasting picturesque landscapes and a favorable climate. Situated at coordinates 30.3165° N and 78.0322° E, the city stands at an elevation of 2100 feet above sea level.

It has been partitioned into six blocks, further subdivided into 767 villages. The research encompassed an area of approximately 3088 km^2, comprising various land cover types such as built-up areas (urbanization), agriculture (low vegetation), water bodies, and forests (high vegetation). Figure 1 shows the entire flow of working culture adopted and Figure 2 depicts the geographical location of the study area (Dehradun).

3.2. Data sets

The Earth Engine public data catalogue hosts a vast simulated collection of commonly used geographic datasets, spanning multi-petabytes. Predominantly, this collection comprises Earth-monitoring remote sensing imagery, including the complete Landsat (Viana et al., 2019) archive, MODIS Archive, and Sentinel data archives. Regular updates are conducted from ongoing missions, amounting to approximately 6000 scenes per 24 hours, with an average latency of around 1440 minutes from scene acquisition time. Users have the flexibility to request the addition of new datasets to the public catalogue or upload their private data via a REST interface, facilitating sharing with other users or groups as needed, utilizing either browser-based or application programming interface (API) methods. The satellite program jointly developed by the USGS and Landsat has provided precise and accurate data since 1972. In Google Earth Engine, Landsat data is available in various forms such as surface reflectance, top of atmospheric corrected reflectance (Fernandino et al., 2018), among others. This facilitates developers in computing vegetation indices, including normalized vegetation index or enhanced vegetation indices. Landsat data directories are organized as subcategories from Landsat 1 to Landsat 9.

Table 1. Resolution and year of availability of Landsat 1–Landsat 5 data sets

Data layer	Source	Pixel size	Band	Year of availability
Landsat 1	USGS	30 m–60 m	B4, B5, B6, B7, QA PIXEL	1972–1978
Landsat 2	USGS	30 m–60 m	B4, B5, B6, B7, QA PIXEL	1975–1982
Landsat 3	USGS	30 m–60 m	B4, B5, B6, B7, QA PIXEL	1978–1983
Landsat 4	USGS	30 m–60 m	B1, B2, B3, B4, QA PIXEL	1982–1992
Landsat 5	USGS	30 m–60 m	B1, B2, B3, B4, QA PIXEL	1984–2013

Table 2. Resolution and year of availability of Landsat 7 data sets

Landsat 7	Source	Resolution pixel size	Band	Year of availability
Surface reflectance	USGS	30 m	SR-B1, SR-B2, SR-B3, SR-B4, SR-B5, SR-B6, SR-B7, SR-Cloud-A, SR-ATMOS-Opacity B1, B2, B3, B4, B5	1999–2022
Top of atmospheric	USGS	30 m–60 m	B6-VCID-1, B6-VCID-2, B7, QA-PIXEL B1, B2, B3, B4, B5	1999–2022
Raw scenes	USGS	30 m–60 m	B6-VCID-1, B6-VCID-2, B7, QA-PIXEL	1999–2022

Data sets Landsat 1 to Landsat 3 have resolution 60 meter and having bands combination Band 4, Band 5, Band 6 and Band 7. Each bands having different feature like wavelength Band 4 have 0.5 to 0.6, Band 5 have 0.6 to 0.7, Band 6 have 0.7 to 0.8 and Band 7 have 0.8 to 1.1. These data sets provides mainly multi spectral scanner images and having approximate scene size is 170 km north-south and 185 km east-west. Landsat 4 and Landsat 5 data set having two types of images Multi spectral scanner and Thematic Mapper. Multi spectral scanner having data of resolution 60 meter with band combination Band 1, Band 2, Band 3 and Band 4 but Thematic Mapper images having different resolution, images of thematic mapper having resolution 30 meter and band combination Band 1, Band 2, Band 3, Band 4, Band 5, Band 6 and Band 7. Table 1 showing short description about Landsat 1 to Landsat 5 with property of Multi Spectral Sensor. Landsat 7 data sets is published in year 1999, having Enhanced Thematic Mapper Plus (ETM+) pictures with 8 bands, which have a spatial resolution of 30 metres for Bands 1 to Band 7. Panchromatic Band 8 has a 15 metre resolution. Band 6 collects both high and low gain for all scenes, while the other bands can only collect one of the two gain settings (high or low) for greater radiometric sensitivity and dynamic range.

Table 3. Resolution and year of availability of Landsat 8 data sets

Landsat 8	Source	Resolution pixel size	Band	Year of availability
Surface reflectance	USGS	30 m	SR-B1, SR-B2, SR-B3, SR-B4, SR-B5, SR-B7	2013 – present date
Top of atmospheric	USGS & Google	15 m–30 m	SR-QA (Aerosol Attribute), B1, B2, B3, B4, B5, B6, B7	2013 – present date
Raw scenes	USGS & Google	15 m–30 m	B8, B9, B10, B11, QA-PIXEL, B1, B2, B3, B4, B5, B6, B7, B8, B9, B10, B11, QA-PIXEL	2013 – present date

Table 4. Resolution and year of availability of Landsat 9 data sets

Landsat 9	Source	Resolution pixel size	Band	Year of availability
Surface reflectance	USGS	30 m	SR-B1, SR-B2, SR-B3, SR-B4, SR-B5, SR-B7	2021 – present date
Top of atmospheric	USGS & Google	15 m–30 m	SR-QA (Aerosol Attribute), B1, B2, B3, B4, B5, B6, B7, B8, B9, B10, B11, QA-PIXEL	2021 – present date
Raw scenes	USGS	15 m–30 m	B1, B2, B3, B4, B5, B6, B7, B8, B9, B10, B11, QA-PIXEL	2021 – present date

The scene measures around 170 kilometres north-south to 183 km east-west. Different categories of Landsat 7 data sets are present in Table 2. The availability of these data sets is from year 1999 to April 2022.

Landsat 8 Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS) images consist of nine spectral bands with a spatial resolution of 30 meters for Bands 1 to 7 and 9. Band 1 (ultra-blue) is useful for coastal and aerosol studies. Band 9 is useful for cirrus cloud detection. The resolution for Band 8 (panchromatic) is 15 meters. Thermal bands 10 and 11 are useful in providing more accurate surface temperatures and are collected at 100 meters. The approximate scene size is 170 km north-south to 183 km east-west. Table 3 illustrate the different data set present in Landsat 8 dictionary. We are using surface reflectance data sets for our observation throughout the year. Total outcome from Tables (1, 2, 3, 4) is that if any user want to do their vegetation analysis or change in vegetation then he/she should use Band 6 and Band 7 if Landsat 1 or Landsat 2 or Landsat 3 data set chosen. Band 3 and Band 4 if Landsat 4 or Landsat 5 data set chosen, for data sets Landsat 7, Landsat 8 and Landsat 9 choose Band 2, Band 3 and Band 4 also known as Red, Green, Blue. Chosen of data sets is also dependent on availability of data in that year, anyone can know availability if these data archive by looking at Table.

Table 5. Resolution and year of availability of Sentinel data sets

Sentinel	Source	Resolution pixel size	Band	Year of availability
Sentinel-1 & SAR GRD	Copernicus	10 m	HH, HV, VH, VV, B1, B2, B3, B4, B5, B6	2014 – present date
Sentinel-2 MSI-1C	Copernicus	10 m–60 m	B7, B8, B8A, B9, B11, B12, QA10, QA20, QA60, B1, B2, B3, B4, B5, B6, B7, B8, B8A, B9, B11	2014 – present date
Sentinel-2 & MSI-2A	Copernicus	10 m–60 m	B12, AOT, WVP, SCL, TCI-R, TCI-G, TCI-B, MSK-CLDPRB, MSK-SNWPRB, QA10, QA20, QA60	2017 – present date
Sentinel-3 & OLCI EFR	Copernicus	300 m	Oa01-radiance to Oa21-radiance quality $_f$ *lags*	2016 – present date

The Copernicus program, overseen by the European Space Agency, provides another dataset collection known as Sentinel. This collection includes Sentinel 1A, Sentinel 1B, Sentinel 2A, Sentinel 2B, Sentinel 3, and Sentinel 5P, each serving specific purposes. Sentinel 1A and Sentinel 1B datasets offer weather radar images, while Sentinel 2A and Sentinel 2B provide high-resolution optical images. Sentinel 3 includes images used for environmental and climate

monitoring, and Sentinel 5P comprises photographs for air quality indexing. Table 5 outlines the different versions of Sentinel datasets, along with their resolutions, band combinations, and years of availability. Sentinel data, with its distinct band combinations and resolutions compared to Landsat, has been available since 2014 for Sentinel 1A and Sentinel 1B, 2017 for Sentinel 1C top of atmospheric reflectance, 2015 for Sentinel 2 surface reflectance, and 2016 for Sentinel 3. Detailed comparisons of band-wise characteristics between these two data types are discussed in (Forkuor et al., 2018).

3.3. Classification algorithm

Classification and Regression Tree: Leo Breiman (Pintelas & Livieris, 2020) dicussed his idea about CART. CART (Classification and Regression Tree) is a type of decision tree algorithm used for both classification and regression tasks. It builds a hierarchical tree structure where each internal node represents a decision based on a feature, and each leaf node represents the final prediction. CART (Ghanem et al., 2021) recursively splits the data into subsets based on the features, aiming to maximize homogeneity within each subset. It predicts the target variable by traversing the tree from the root to the appropriate leaf node based on the input features. CART is simple, interpretable, and handles both numerical and categorical data, but can be prone to overfitting. CART is working on the principal of Gini Index and Gini Gain. The Gini index measures the divergences between the probability distributions of the target attributes values and is defined as

$$Gini = 1 - \sum_{i=1}^{c} Pi^2, \quad (1)$$

where c – number of class; P – probability of an object being classified.

Random Forest: Random Forest is an ensemble learning algorithm used for classification and regression tasks. It constructs multiple decision trees during training and outputs the mode of the classes (for classification) or the average prediction (for regression) of individual trees. Random Forest introduces randomness in two key ways: by using a random subset of features for each tree and by training each tree on a random subset of the training data with replacement (bootstrapping). This randomness helps reduce overfitting and improves generalization performance. During prediction, the final outcome is determined by aggregating the predictions of all trees in the forest. Random Forest is known for its high accuracy, robustness to noise, and ability to handle large datasets with high dimensionality.

Gradient Tree Boosting: Gradient Tree Boosting (Friedman, 2001), or Gradient Boosting Machine (GBM), is an ensemble learning technique that sequentially builds a series of decision trees to correct the errors of the preceding ones. Each tree is constructed to predict the residuals of the ensemble's current prediction, with the overall goal of minimizing a chosen loss function. Gradient descent optimization (Srivastava & Ahmad, 2016) is used to fit each tree to the residuals, and a small learning rate is typically applied to prevent over fitting and improve generalization. GBM is highly effective in capturing complex relationships in data and handling heterogeneous data types, making it a popular choice for regression and classification tasks. However, it requires careful tuning of hyper parameters (Nooni et al., 2014) to avoid over fitting and achieve optimal performance.

Support Vector Machine: In 1963, Vladimir Vapnik and Alexey Chervonenkis developed the first SVM model. Support Vector Machine (SVM) is a powerful supervised learning algorithm used for both classification and regression tasks. In classification, SVM aims to find the optimal hyperplane that separates data points into different classes with the largest possible margin. It achieves this by identifying support vectors (Evgeniou & Pontil, 2014; El Morr et al., 2022; Osuna et al., 1999), which are the data points closest to the decision boundary. In cases where a linear boundary cannot separate the classes effectively, SVM can use a kernel trick to map the data into a higher-dimensional space where a linear boundary becomes possible. SVM is effective in handling high-dimensional data and is known for its ability to generalize well, especially in cases with small to medium-sized datasets. However, SVM's performance can be sensitive to the choice of kernel and regularization parameters, and it may struggle with large datasets due to its computational complexity (Nooni et al., 2014).

Accuracy: Analysis of any data or information is a very important phase of data processing, and minor mistakes in this analysis may affect large changes in results, so data processing must be highly accurate and up to date. For Accuracy measurement many validation and verification points are generalized like how accurate data is collected and also how accurate data is processed. So here we take 30% data points as validation points, and the other 70% data points are used for analysis purposes. Accuracy is calculated by using Confusion Matrix and Kappa Coefficient. Accuracy, Precision, Recall and F1 Score is calculated by following formula

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}; \quad (2)$$

$$Precision = \frac{TP}{TP + FP}; \quad (3)$$

$$Recall = \frac{TP}{TP + FN}; \quad (4)$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} = \frac{2 \times TP}{2 \times TP + FP + FN}, \quad (5)$$

where TP – True Positive; TN – True Negative; FP – False Positive; FN – False Negative.

The Kappa coefficient (Pontius & Millones, 2011), also known as Cohen's Kappa, is a statistical measure

of inter-rater agreement for categorical items. It measures the extent of agreement between two raters or observers beyond what would be expected by chance alone. Kappa values range from –1 to 1, with 1 indicating perfect agreement, 0 indicating agreement equivalent to chance, and negative values indicating agreement worse than chance. Kappa is commonly used in fields such as psychology, medicine, and linguistics to assess the reliability of categorical data coding. Kappa value is calculated by:

$$Kappa = \frac{Po - Pe}{1 - Pe}, \quad (6)$$

where *Po* – Observed proportional agreement; *Pe* – Expected proportional agreement.

4. Result

As discussed previously in Section 2, the study area has been subdivided into several classes based on different land cover types. Any unusual changes in one land cover area, such as forests being affected by fire or heavy rainfall, can significantly impact other land cover types like urban, water, and agriculture. Unplanned urban growth, as noted in (Gupta, 2015) can also disturb other land cover types. Our research indicates that urban growth was 2.68% (919,000 to 943,000) between 2020 and 2021 and 2.55% (943,000 to 967,000) between 2021 and 2022. Such unplanned growth can disrupt the balance of other land cover classes. We divided the total area of 3,088 sq km into four main land cover types: urban, forest, water, and agriculture. Further subdivisions include residential, institutional, built-up areas, and parking for urban land; plantation, cultivation, and other farmer lands for agriculture; and dense forest and open forest for forest land. Water land cover is divided into ponds, lakes, and rivers. We collected 536 data points for urban land cover, 506 for forest, 505 for water, and 540 for agriculture shown in Table 6.

Table 6. Training points and legend of landcover classes

Land cover ID	Land cover class	Number of samples	Color
1	Urban	536	0000ff
2	Forest	506	008000
3	Water	505	ff0000
4	Agriculture	540	ffff00

For classification purposes, we are considering four classification algorithms on two datasets: Sentinel-2 (multispectral instrument) and Landsat 8 (surface reflectance). Urban, forest, water, and agriculture classes are represented by Blue, Green Red and Yellow color, respectively. The results include different outcomes generated by each pair of months using both datasets. we present the comparative analysis of Sentinel-2 and Landsat 8 satellite datasets over different seasons of the year. The study aims to evaluate the performance of these two widely used remote sensing platforms in capturing seasonal variations in land cover and environmental dynamics. By leveraging the multi-temporal imagery from both sensors, we examine their spectral characteristics, spatial resolutions, and temporal coverage to assess their suitability for various applications, including land cover classification, vegetation monitoring, and environmental assessments.

Figure 3 contain the data generated over sentinel dataset after processing classification algorithm and Figure 4 contain data generated by Landsat 8 data sets, by the

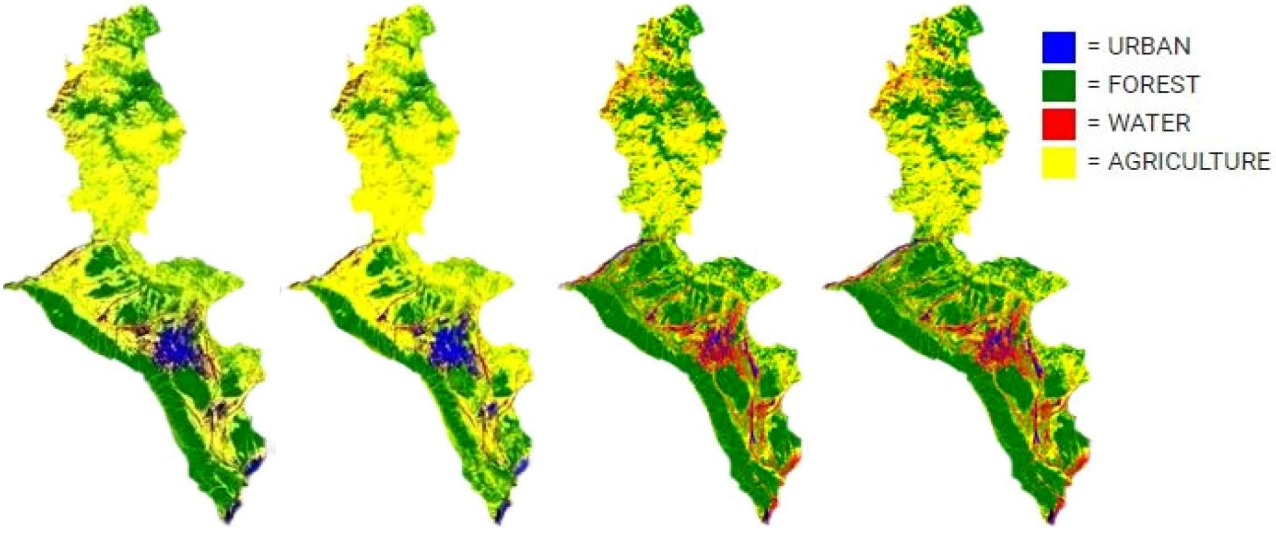


Figure 3. Collection of output data generated by four classification algorithm (Classification and Regression Tree, Random Forest, Gradient Tree Boost and Support Vector Machine) for the month of January and February 2021 on data sets Sentinel-2 and resolution 30 m

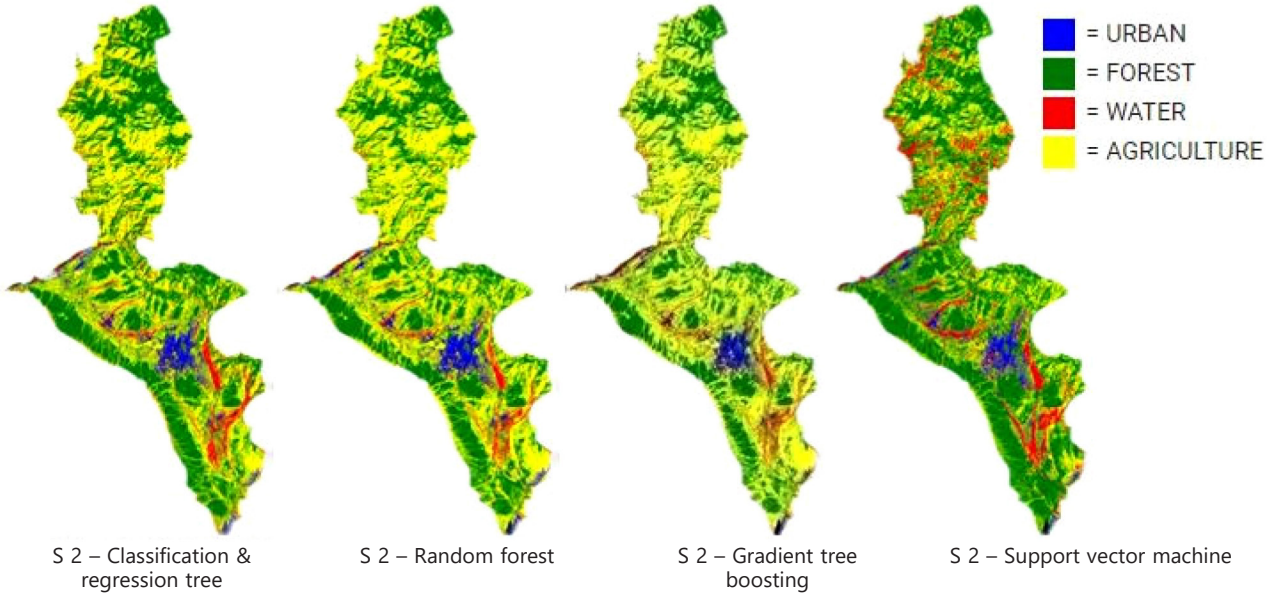


Figure 4. Collection of output data generated by four classification algorithm (Classification and Regression Tree, Random Forest, Gradient Tree Boost and Support Vector Machine) for the month of January and February 2021 on data sets Landsat 8 and resolution 30 m

processing of these two data sets we found urban area capture 169 sq km, forest area 1274 sq km vegetation area 1526 sq km and water area 119 sq km. Here urban, forest and vegetation are having some overlap data these data are looking as different colour combination which we not mention in our research, so for finding the accuracy we calculate confusion matrix of our algorithms over both data sets sentinel and Landsat respectively and CART have 89.24% and 93.52% by confusion matrix and 85.64% and 91.36% by kappa, Random forest have 91.45% and 95.86% by confusion matrix and 88.59% and 94.48% by kappa, GTB have 87.71% and 95.33% by confusion matrix and 83.58% and 93.77% by kappa, SVM have 84.96% and 73.54% by confusion matrix and 79.99% and 76.28% by kappa respectively.

We further calculate producer and consumer accuracy of each land cover class i.e. urban, forest water and agriculture for the problem of overlapping of data in both

Table 7. Accuracy of each classification algorithm for January and February month

ALGO	CLASS	S2(SENTINAL DATA SET)				L8(LANDSAT 8 DATA SET)			
		TOTAL ACCURACY	CONSUMER ACCURACY	PRODUCER ACCURACY	KAPPA ACCURACY	TOTAL ACCURACY	CONSUMER ACCURACY	PRODUCER ACCURACY	KAPPA ACCURACY
CART	URBAN	0.8924	0.8954	0.8838	0.8564	0.9352	0.9215	0.9333	0.9136
	FOREST		0.8947	0.9216			0.9281	0.9403	
	WATER		0.8758	0.8639			0.9415	0.9354	
	AGRICULTURE		0.9018	0.8963			0.9514	0.9315	
RANDOM FOREST	URBAN	0.9145	0.9079	0.9548	0.8859	0.9586	0.9532	0.9659	0.9448
	FOREST		0.9058	0.9277			0.9523	0.9756	
	WATER		0.9469	0.8503			0.9738	0.9494	
	AGRICULTURE		0.9041	0.9207			0.9559	0.9441	
GRADIENT TREE BOOST	URBAN	0.8771	0.8281	0.9244	0.8358	0.9533	0.9613	0.9942	0.9377
	FOREST		0.8881	0.9407			0.9411	0.9171	
	WATER		0.8911	0.8136			0.9686	0.9807	
	AGRICULTURE		0.9172	0.8367			0.9403	0.9161	
SUPPORT VECTOR MACHINE	URBAN	0.8496	0.8284	0.9032	0.7999	0.7354	0.8125	0.8924	0.7628
	FOREST		0.8245	0.8493			0.8041	0.7777	
	WATER		0.9236	0.8231			0.7712	0.7065	
	AGRICULTURE		0.8382	0.8231			0.7125	0.7625	

data sets, consumer accuracy for Sentinel 2 data sets is 89.54%, 89.47%, 87.58% and 90.18% and for Landsat 8 data sets 92.15%, 92.81%, 94.15% and 95.14%, producer accuracy for Sentinel 2 data sets is 88.36%, 92.16%, 86.39% and 89.63% and for Landsat 8 data sets 93.33%, 94.03%, 93.54% and 93.15% respectively for each land cover classes. Here we write accuracy of CART (Classification and Regression Tree) with all landcover classes. Accuracies of all machine learning for January and February month with detailed land cover classes is listed in Table 7.

Figure 5 and Figure 6 contain the output data generated over the sentinel and Landsat 8 datasets respectively, by processing of these two data sets we found there is a little bit of change in all land cover classes if we compare Figure 3 and Figure 4 by Figure 5 and Figure 6 then also it easily concluded that change occur in some land cover classes, forest area is decreases and water area is increasing, when we go behind reason of this we found flood situation is occurred in some area of Dehradun total output generated by analysis of

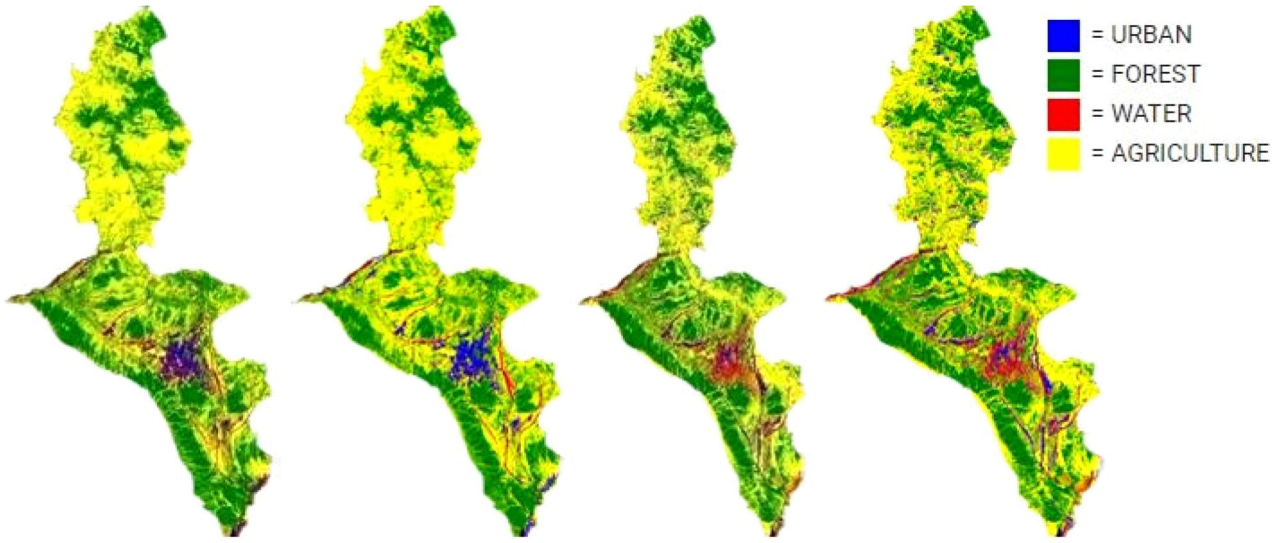


Figure 5. Collection of output data generated by four classification algorithm (Classification and Regression Tree, Random Forest, Gradient Tree Boost and Support Vector Machine) for the month of March and April 2021 on data sets Sentinel-2 and resolution 30 m

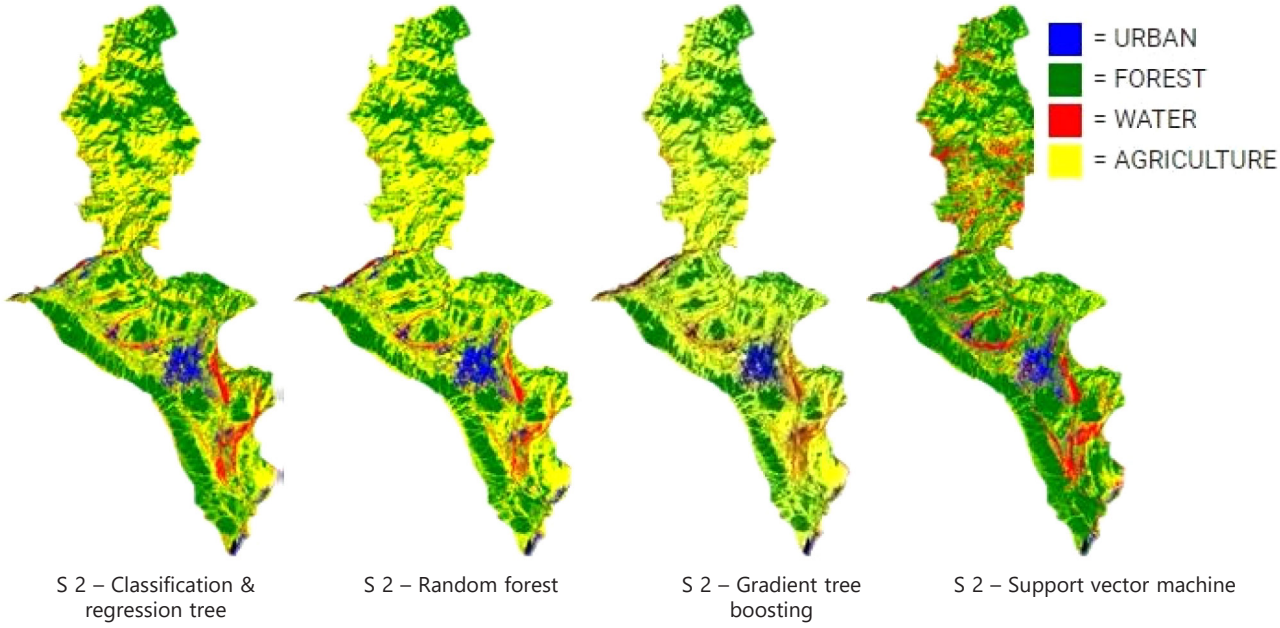


Figure 6. Collection of output data generated by four classification algorithm (Classification and Regression Tree, Random Forest, Gradient Tree Boost and Support Vector Machine) for the months of March and April 2021 on data sets Landsat 8 and resolution 30 m

data over March and April month is urban area 174 sq km, forest area 1061 sq km vegetation area 1629 sq km and water area 223 sq km. The problem of overlapping of data points is also available here, therefore accuracy estimation of each classification algorithm over all land-cover classes is necessary.

So to find the accuracy, we calculate the confusion matrix of our algorithms over both data sets sentinel and Landsat respectively and CART have 85.06% and 93.27% by confusion matrix and 80.08% and 91.04% by kappa, Random forest has 87.12% and 92.03% by confusion matrix and 82.82% and 89.35% by kappa, GTB have 89.68% and 91.58% by confusion matrix and 86.22% and 88.77% by kappa, SVM have 82.12% and 75.39% by confusion matrix and 76.13% and 74.98% by kappa respectively. We further calculate producer and consumer accuracy of each land

Table 8. Accuracy of each classification algorithm for March and April month

ALGO	CLASS	S2(SENTINEL-2 DATA SET)				L8(LANDSAT 8 DATA SET)			
		TOTAL ACCURACY	CONSUMER ACCURACY	PRODUCER ACCURACY	KAPPA ACCURACY	TOTAL ACCURACY	CONSUMER ACCURACY	PRODUCER ACCURACY	KAPPA ACCURACY
CART	URBAN	0.8506	0.8902	0.8342	0.808	0.9327	0.9387	0.9387	0.9104
	FOREST		0.8303	0.8338			0.9303	0.9245	
	WATER		0.8497	0.8448			0.9299	0.9419	
	AGRICULTURE		0.8544	0.8653			0.9329	0.9272	
RANDOM FOREST	URBAN	0.8712	0.8779	0.8628	0.8282	0.9203	0.9554	0.9555	0.8935
	FOREST		0.8511	0.9225			0.895	0.9035	
	WATER		0.8666	0.8218			0.9481	0.9135	
	AGRICULTURE		0.8903	0.8846			0.8894	0.9086	
GRADIENT TREE BOOST	URBAN	0.8968	0.9386	0.9053	0.8622	0.9158	0.9454	0.9341	0.8877
	FOREST		0.8818	0.8875			0.906	0.9121	
	WATER		0.8807	0.8866			0.9271	0.9271	
	AGRICULTURE		0.8852	0.9055			0.8823	0.8881	
SUPPORT VECTOR MACHINE	URBAN	0.8212	0.8253	0.8914	0.7613	0.7589	0.8126	0.9263	0.7498
	FOREST		0.7619	0.929			0.7088	0.8175	
	WATER		0.8417	0.7643			0.7723	0.7891	
	AGRICULTURE		0.879	0.6987			0.7231	0.6784	

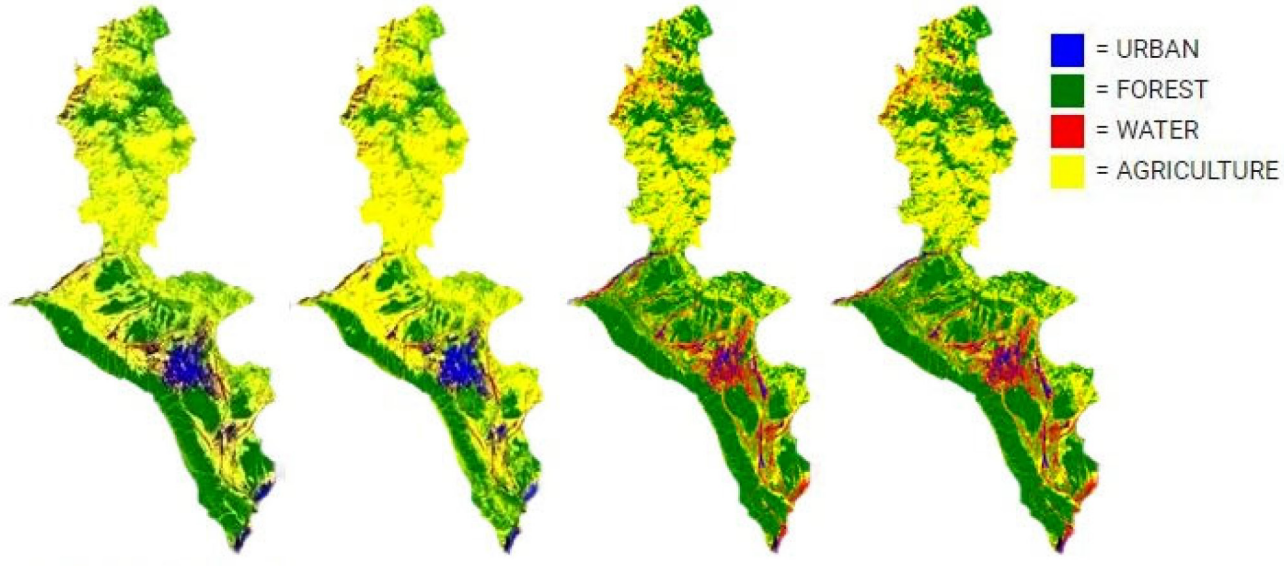


Figure 7. Collection of output data generated by four classification algorithm (Classification and Regression Tree, Random Forest, Gradient Tree Boost and Support Vector Machine) for the month of May and June 2021 on data sets Sentinel-2 and resolution 30 m

cover class i.e. urban, forest water and agriculture for the problem of overlapping of data in both data sets ,consumer accuracy for Sentinel 2 data sets is 89.02%, 89.03%, 84.97% and 85.44% and for Landsat 8 data sets 93.37%, 93.03%, 92.99% and 93.39%, producer accuracy for Sentinel 2 data sets is 83.42%, 88.38%, 84.48% and 89.53% and for Landsat 8 data sets 93.87%, 92.42%, 94.19% and 92.72% respectively for each land cover classes. Here we write the accuracy of CART (Classification and Regression Tree) with all land cover classes. Accuracies of all machine learning for March and April month with detailed land cover classes are listed in Table 8.

On comparing output generated by both data sets with previous output we found some landcover classes changes landcover class urban have minor changes but water and agriculture have large gap by previous analysis if we focus on Figure 7 output generated by Sentinel S2 data sets and Figure 8 output generated by Landsat 8 data sets then it is clearly visible on upper area of Dehradun where in Figures 5 and 6 forest landcover class is present now water class is exist, by the processing of these two data sets we found urban area capture 192 sq km, forest area 899 sq km vegetation area 1540 sq km and water area 462 sq km.

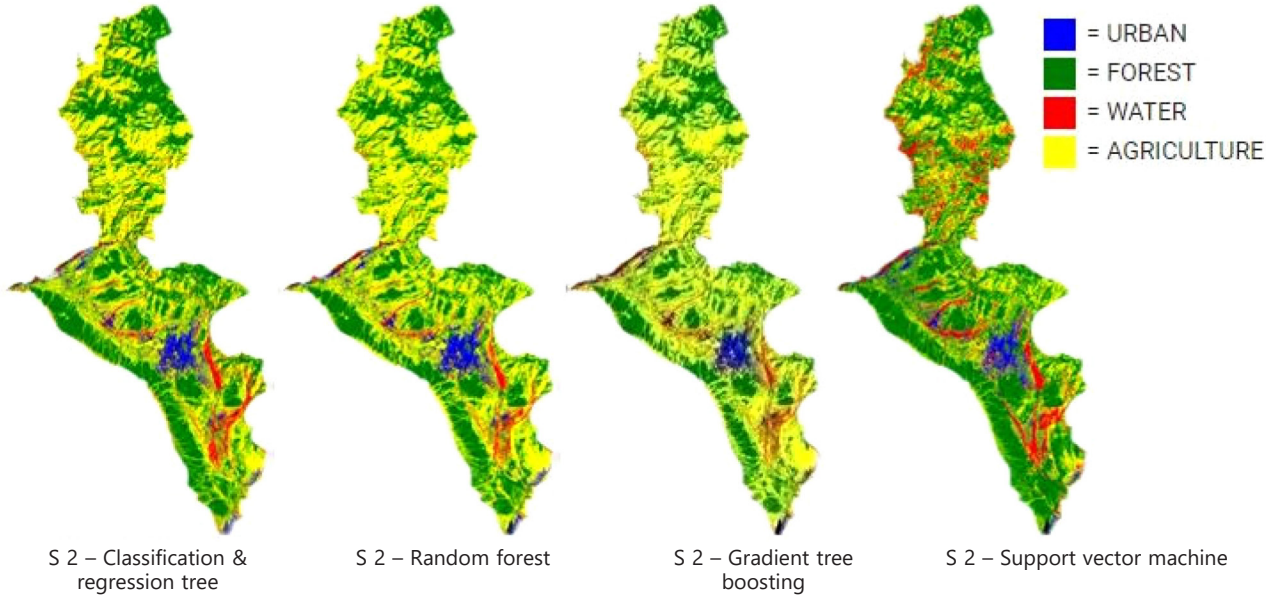


Figure 8. Collection of output data generated by four classification algorithm (Classification and Regression Tree, Random Forest, Gradient Tree Boost and Support Vector Machine) for the month of May and June 2021 on data sets Landsat 8 and resolution 30 m

Table 9. Accuracy of each classification algorithm for month May and June month

ALGO	CLASS	S2(SENTINEL-2 DATA SET)				L8(LANDSAT 8 DATA SET)			
		TOTAL ACCURACY	CONSUMER ACCURACY	PRODUCER ACCURACY	KAPPA ACCURACY	TOTAL ACCURACY	CONSUMER ACCURACY	PRODUCER ACCURACY	KAPPA ACCURACY
CART	URBAN	0.9082	0.916	0.8791	0.8773	0.9102	0.9375	0.9099	0.8802
	FOREST		0.8875	0.9349			0.9215	0.898	
	WATER		0.8888	0.9189			0.9271	0.9454	
	AGRICULTURE		0.9375	0.9016			0.8596	0.8909	
RANDOM FOREST	URBAN	0.9145	0.9373	0.9127	0.8858	0.9348	0.98	0.9545	0.913
	FOREST		0.8666	0.9407			0.9022	0.9631	
	WATER		0.9172	0.8986			0.9319	0.9319	
	AGRICULTURE		0.9378	0.9071			0.9303	0.8909	
GRA-DIENT TREE BOOST	URBAN	0.9229	0.8933	0.9503	0.8959	0.9514	0.9677	0.989	0.9349
	FOREST		0.952	0.9144			0.939	0.9506	
	WATER		0.9408	0.8784			0.9777	0.923	
	AGRICULTURE		0.9024	0.9548			0.9215	0.9337	
SUPPORT VECTOR MACHINE	URBAN	0.742	0.6433	0.6778	0.7127	0.7329	0.6939	0.8903	0.7752
	FOREST		0.7595	0.9144			0.7725	0.8703	
	WATER		0.7151	0.8301			0.7124	0.7254	
	AGRICULTURE		0.8833	0.6792			0.6624	0.6725	

Here all classes have some overlapped data and these data are present as different colour combination which we not mention in our research, therefore accuracy calculation is a necessary part of our research analysis.

So for finding the accuracy we calculate confusion matrix of our algorithms over both data sets sentinel and Landsat respectively and CART have 92.32% and 91.02% by confusion matrix and 87.73% and 89.02% by kappa, Random forest have 91.45% and 93.48% by confusion matrix and 88.58% and 91.30% by kappa, GTB have 92.29% and 95.14% by confusion matrix and 89.59% and 93.39% by kappa, SVM have 74.20% and 73.29% by confusion matrix and 71.27% and 77.52% by kappa respectively. We further calculate producer and consumer accuracy of each land cover class i.e. urban, forest water and agriculture for the problem of overlapping of data in both data sets. Detailed accuracy with all landcover classes for all machine learning of May and June month with is listed in Table 9.

We all know that Landsat datasets have one additional property over sentinel that it have band of infrared sensor also and pros of these bands are also visible in output data Figure 9, output of sentinel by four machine learning it clearly visible to any non programing candidate also that here is lot of data are overlapping each other due

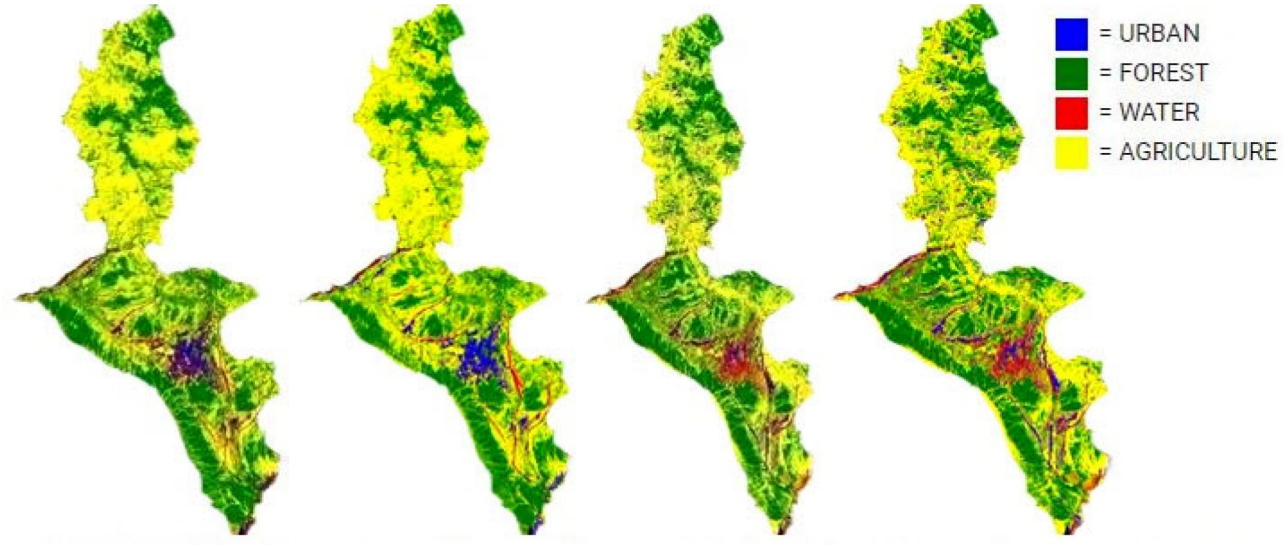


Figure 9. Collection of output data generated by four classification algorithm (Classification and Regression Tree, Random Forest, Gradient Tree Boost and Support Vector Machine) for the month of July and August 2021 over the data generated by Sentinel-2 and resolution 30 m

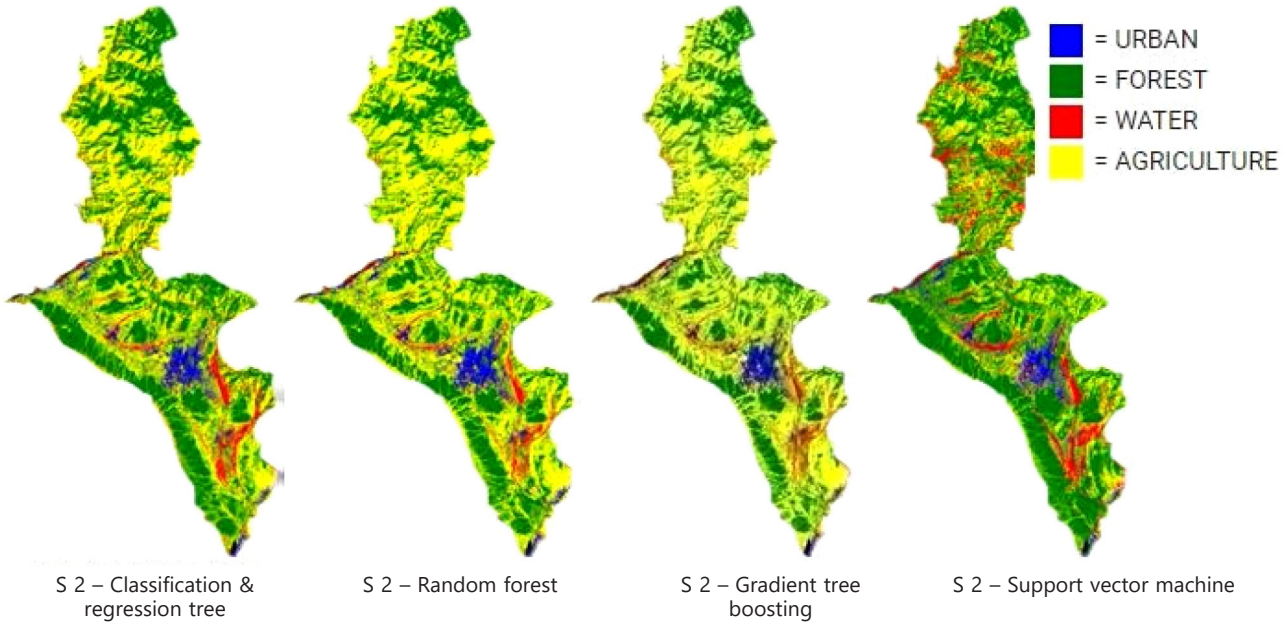


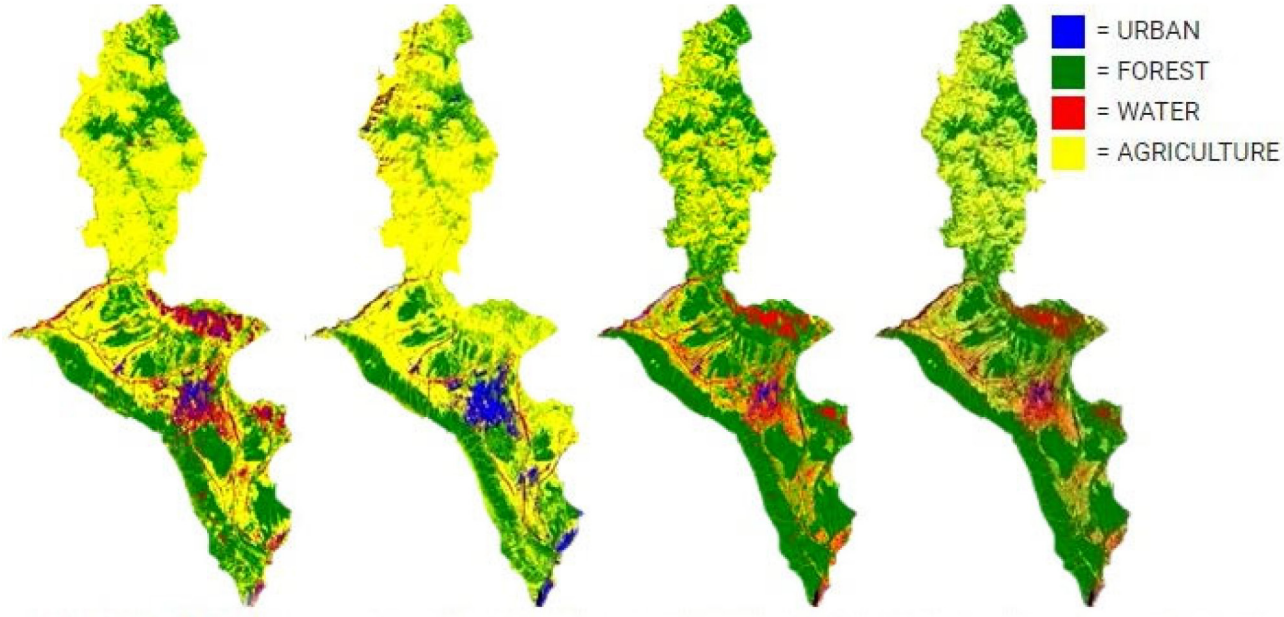
Figure 10. Collection of output data generated by four classification algorithm (Classification and Regression Tree, Random Forest, Gradient Tree Boost and Support Vector Machine) for the month of July and August 2021 on data sets Landsat 8 and resolution 30 m

to this it provide a different output than Landsat 8 data sets have in Figure 10, it have less overlapped data and high accuracy also in compare to sentinel data, by the processing of these two data sets we found urban area capture 206 sq km, forest area 1040 sq km vegetation area 1452 sq km and water area 390 sq km. Here urban, forest and vegetation are having some overlap data these data are looking as different colour combination which we not mention in our research.

So for finding the accuracy we calculate accuracy of our algorithms over both data sets sentinel and Landsat respectively and find GTB have maximum accuracy 89.37% by confusion matrix and 85.79% by kappa in all classification algorithm on sentinel data sets but on Landsat

Table 10. Accuracy of each classification algorithm for month July and August month

ALGO	CLASS	S2(SENTINEL-2 DATA SET)				L8(LANDSAT 8 DATA SET)			
		TOTAL ACCURACY	CONSUMER ACCURACY	PRODUCER ACCURACY	KAPPA ACCURACY	TOTAL ACCURACY	CONSUMER ACCURACY	PRODUCER ACCURACY	KAPPA ACCURACY
CART	URBAN	0.8304	0.8715	0.8719	0.7737	0.9388	0.951	0.951	0.9185
	FOREST		0.8488	0.8202			0.9591	0.9144	
	WATER		0.7828	0.7579			0.9787	0.9324	
	AGRICULTURE		0.8129	0.875			0.8919	0.9589	
RANDOM FOREST	URBAN	0.8849	0.9041	0.9207	0.8462	0.9553	0.9673	0.9834	0.9403
	FOREST		0.9152	0.9101			0.9156	0.95	
	WATER		0.86	0.8216			0.9788	0.9788	
	AGRICULTURE		0.8523	0.8819			0.962	0.9101	
GRA-DIENT BOOST	URBAN	0.8937	0.9444	0.9212	0.8579	0.9379	0.9451	0.9748	0.9172
	FOREST		0.8296	0.8175			0.9006	0.9731	
	WATER		0.9636	0.95			0.9347	0.947	
	AGRICULTURE		0.9285	0.8764			0.9776	0.8562	
SUPPORT VECTOR MACHINE	URBAN	0.7071	0.7641	0.9067	0.6945	0.7649	0.8052	0.9107	0.7152
	FOREST		0.7152	0.7845			0.6833	0.7935	
	WATER		0.709	0.6752			0.6458	0.7152	
	AGRICULTURE		0.7066	0.7092			0.9666	0.6408	



S 2 – Classification & regression tree | S 2 – Random forest | S 2 – Gradient tree boosting | S 2 – Support vector machine

Figure 11. Collection of output data generated by four classification algorithm (Classification and Regression Tree, Random Forest, Gradient Tree Boost and Support Vector Machine) for the month of September and October 2021 on data sets Sentinel-2 and resolution 30 m

datasets random forest archive maximum accuracy 95.23% by confusion matrix and 94.03 and by kappa. Accuracies of all landcover class and machine learning algorithms for July and August month is listed in Table 10.

When we process the data of September and October on both datasets sentinel and Landsat by all machine learning algorithm, we found some changes Figure 11 is output generated by Sentinel 2 data sets and Figure 12 shows output generated by Landsat datasets, on doing more research for finding these changes, we found these season are mostly rainy in study area so increase in forest water and agriculture is normal, but due to rainy season growth in small vegetation is very high so vegetation class have more data point in comparison these two, after processing classification algorithm we found urban area capture 152 sq km, forest area 1117 sq km vegetation area 1579 sq km and water area 240 sq km. Here urban, forest and vegetation are having some overlap data these data are looking as different colour combination which we not mention in our research.

Accuracy is calculated over both data sets sentinel and Landsat respectively and we find random forest perform

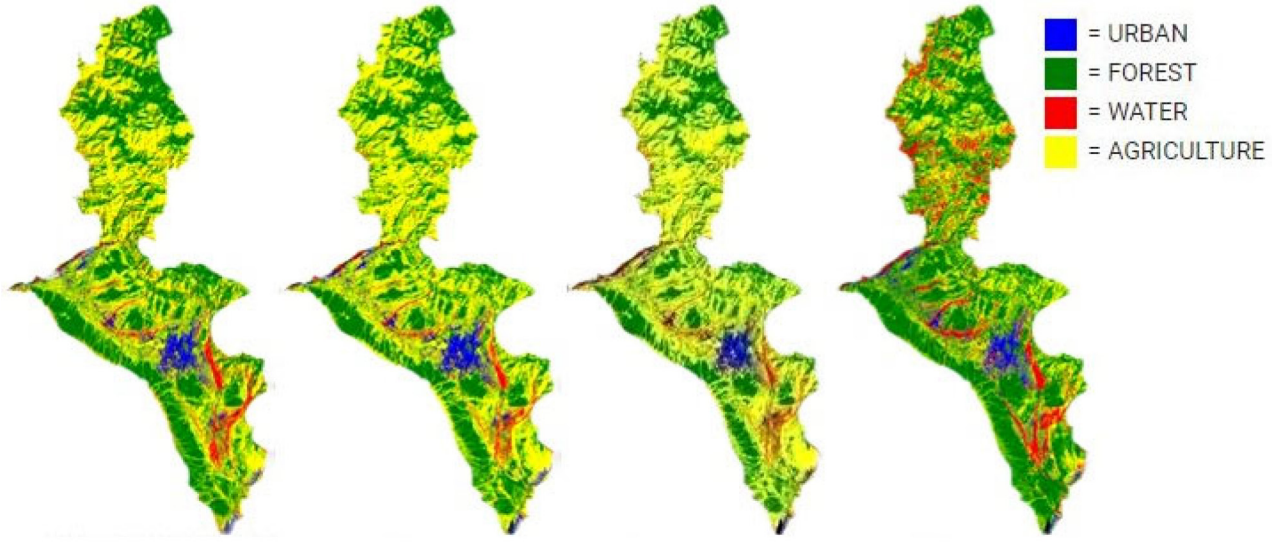


Figure 12. Collection of output data generated by four classification algorithm (Classification and Regression Tree, Random Forest, Gradient Tree Boost and Support Vector Machine) for the month of September and October 2021 on data sets Landsat 8 and resolution 30 m

Table 11. Accuracy of each classification algorithm for September and October month

ALGO	CLASS	S2(SENTINEL-2 DATA SET)				L8(LANDSAT 8 DATA SET)			
		TOTAL ACCURACY	CONSUMER ACCURACY	PRODUCER ACCURACY	KAPPA ACCURACY	TOTAL ACCURACY	CONSUMER ACCURACY	PRODUCER ACCURACY	KAPPA ACCURACY
CART	URBAN	0.8662	0.8875	0.8452	0.8216	0.9147	0.932	0.9378	0.8861
	FOREST		0.8924	0.9034			0.9097	0.9097	
	WATER		0.7882	0.8481			0.922	0.893	
	AGRICULTURE		0.9032	0.8695			0.8944	0.9171	
RANDOM FOREST	URBAN	0.9004	0.9146	0.8928	0.8673	0.9383	0.9696	0.9638	0.9175
	FOREST		0.8771	0.9615			0.9308	0.9426	
	WATER		0.8392	0.8924			0.9253	0.9185	
	AGRICULTURE		0.9857	0.8571			0.924	0.924	
GRA-DIENT TREE BOOST	URBAN	0.8902	0.8934	0.8934	0.8534	0.9033	0.8786	0.9212	0.871
	FOREST		0.8802	0.9363			0.9133	0.9013	
	WATER		0.85	0.8152			0.9225	0.8614	
	AGRICULTURE		0.9298	0.9085			0.9019	0.9324	
SUPPORT VECTOR MACHINE	URBAN	0.7619	0.9205	0.8663	0.7014	0.7534	0.6623	0.8584	0.671
	FOREST		0.675	0.82			0.7668	0.8774	
	WATER		0.7597	0.7852			0.8105	0.6135	
	AGRICULTURE		0.7722	0.6415			0.8325	0.6629	

better over each scenario datasets it have 90.04% by confusion matrix and 86.73% by kappa on Sentinel 2 data sets and 93.83% by confusion matrix and 91.75% by kappa. Accuracies of all landcover class and machine learning algorithms for September and October Month is listed in Table 11.

Figure 13 contain the data generated over sentinel dataset after processing classification algorithm and Figure 14 contain data generated by Landsat 8 data sets, by the processing of these two data sets we found urban area capture 187 sq km, forest area 1174 sq km vegetation area 1442 sq km and water area 285 sq km output data generated by these two datasets (Figures 13 and 14) have some different colours which we not mentioned in our analysis, so we calculate accuracy of each land cover class for all machine learning algorithm, for both data sets Sentinel 2 and Landsat 8, GTB perform better 88.46% and 89.32% for confusion matrix and 84.60% and 89.16% for kappa respectively.

Accuracies for all landcover class and machine learning algorithms for November and December month is listed in Table 12. Here we found Gradient Tree Boosting is perform better over both data sets.

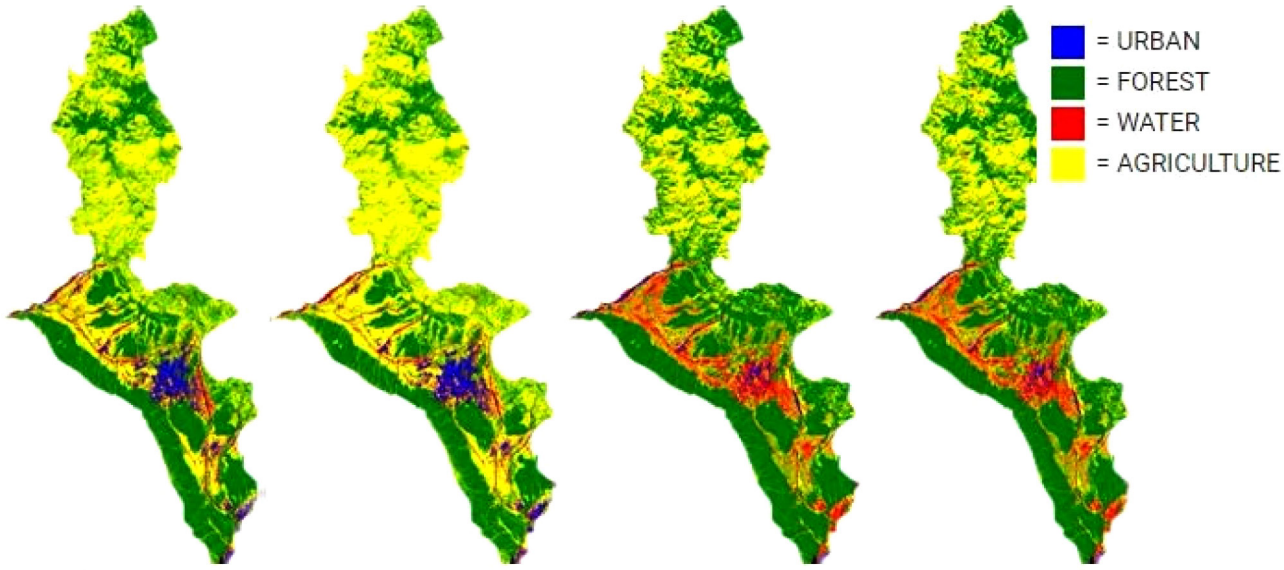


Figure 13. Collection of output data generated by four classification algorithm (Classification and Regression Tree, Random Forest, Gradient Tree Boost and Support Vector Machine) for the month of November and December 2021 on data sets Sentinel-2 and resolution 30 m

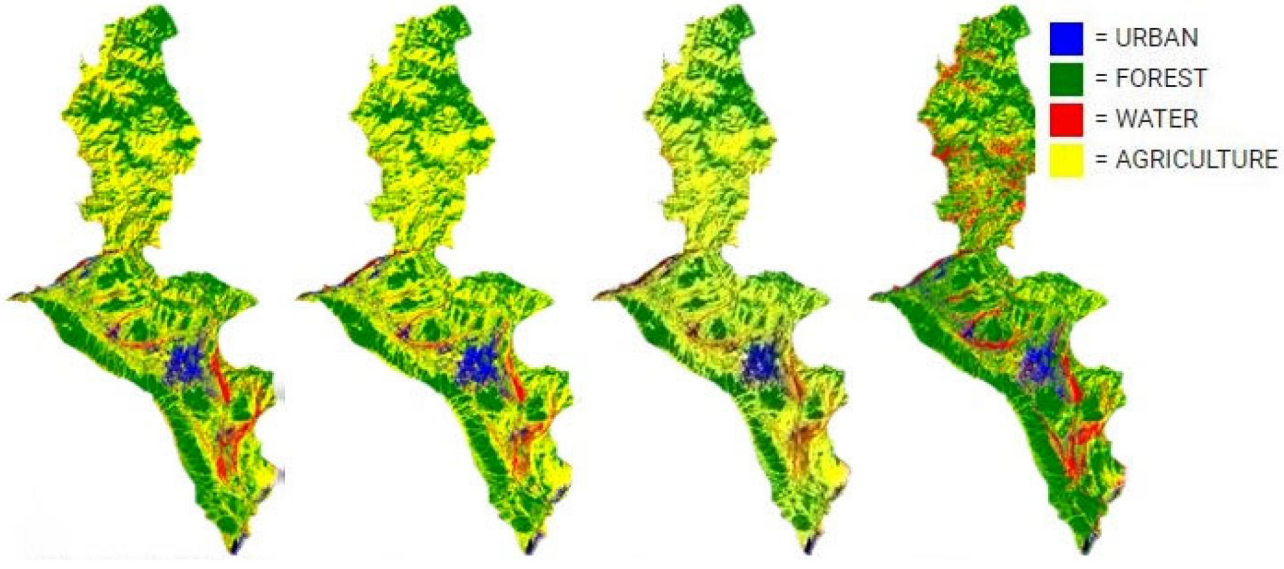


Figure 14. Collection of output data generated by four classification algorithm (Classification and Regression Tree, Random Forest, Gradient Tree Boost and Support Vector Machine) for the month of November and December 2021 on data sets Landsat 8 and resolution 30 m

Table 12. Accuracy of each classification algorithm for November and December month

ALGO	CLASS	S2(SENTINEL-2 DATA SET)				L8(LANDSAT 8 DATA SET)			
		TOTAL ACCURACY	CONSUMER ACCURACY	PRODUCER ACCURACY	KAPPA ACCURACY	TOTAL ACCURACY	CONSUMER ACCURACY	PRODUCER ACCURACY	KAPPA ACCURACY
CART	URBAN	0.8734	0.8915	0.9308	0.8309	0.8984	0.9438	0.923	0.8643
	FOREST		0.9102	0.8606			0.9139	0.9251	
	WATER		0.8345	0.7891			0.8407	0.8461	
	AGRICULTURE		0.8556	0.9039			0.887	0.8971	
RANDOM FOREST	URBAN	0.8641	0.8392	0.8867	0.8185	0.8932	0.9036	0.9515	0.8576
	FOREST		0.9156	0.9212			0.929	0.9385	
	WATER		0.8102	0.7551			0.9103	0.8	
	AGRICULTURE		0.8813	0.8813			0.8289	0.8811	
GRA-DIENT TREE BOOST	URBAN	0.8846	0.8771	0.8928	0.846	0.9187	0.9559	0.95	0.8916
	FOREST		0.9189	0.8774			0.9333	0.9523	
	WATER		0.8333	0.8445			0.855	0.8805	
	AGRICULTURE		0.9085	0.9197			0.9236	0.8866	
SUPPORT VECTOR MACHINE	URBAN	0.7901	0.779	0.8427	0.7194	0.7604	0.7951	0.9106	0.6616
	FOREST		0.7566	0.8666			0.7794	0.8102	
	WATER		0.7966	0.6394			0.6673	0.6715	
	AGRICULTURE		0.8343	0.7966			0.7289	0.5925	

After analysing all season we found that lots of change happen over the study area and classification algorithm like random forest and GTB perform better most of the season in some season CART also having high accuracy but on comparing these three algorithm with support vector machine we always found support vector machine have large of overlapped data so its have less accuracy.

5. Conclusions

The purpose of this study was to analyze changes in different land cover classes in each month set of 2021 in the Dehradun study area and compare the two existing data sets over four machine learning algorithms. We find here landcover classes are shifted from one to another from sets of months due to human conduct activity or by natural activity. Figures (3–15) shows the changes that happened in each landcover class. The classification performance index easily shows the accuracy as well as wellness of each classification algorithm. Output describes that each data sets have some unique feature and behaves according to that feature with different classification algorithms. Landsat (USGS) and sentinel (COPERNICUS) are two different data set providers and provide images of different behavior like Landsat images have thermal infrared sensors but sentinel do not have these bands. In most cases (Gradient Tree Boost) GTB, Classification and regression tree (CART), and Random Forest (RF) performed slightly better than Support Vector Machine (SVM), compared to SVM, decision tree based algorithms are better at handling co-linearity and categorical data, it construct hyper-rectangles in the input space to solve the problem, SVM uses the kernel method to address non-linear problems. Differences in CART, GTB, and RF accuracy estimates were generally statistically insignificant; accuracy of multispectral instrument level (a variation of S2) and surface reflectance (a variation of L8) is almost the same in all the season while the classification algorithm which uses kernel method to solve the nonlinear problem having higher accuracy for multispectral data sets.

There are the following space are left for future research also: We have covered only four land cover areas and finds overlapped data after classification, especially in forest and agriculture land cover areas. So in future, we need to do close observation on these land cover areas by using different vegetation analysis techniques. If any user do their analysis for long term with above classification approach then they don't need to to do it on different classification algorithm they just need to select a classification model and data set with better accuracy and do their analysis for long term changes. User can also select better classification model from Tables (7–12) and perform their analysis for change in single land cover class also.

References

Benediktsson, J. A., Chanussot, J., & Fauvel, M. (2007). Multiple classifier systems in remote sensing: From basics to recent developments. In M. Haindl, J. Kittler, & F. Roli (Eds.), *Lecture notes in computer science: Vol. 4472. Multiple classifier systems* (pp. 501–512). Springer. https://doi.org/10.1007/978-3-540-72523-7_50

Campos-Taberner, M., Moreno-Martínez, Á., García-Haro, F. J., Camps-Valls, G., Robinson, N. P., Kattge, J., & Running, S. W. (2018). Global estimation of biophysical variables from Google Earth Engine platform. *Remote Sensing*, *10*(8), Article 1167. https://doi.org/10.3390/rs10081167

El Morr, C., Jammal, M., Ali-Hassan, H., & El-Hallak, W. (2022). Support vector machine. In *International series in operations*

research and management science: Vol. 334. Machine learning for practical decision making (pp. 385–411). Springer. https://doi.org/10.1007/978-3-031-16990-8_13

Evgeniou, T., & Pontil, M. (2014). Support vector machines: Theory and applications. In G. Paliouras, V. Karkaletsis, & C. D. Spyropoulos (Eds.), *Lecture notes in computer science: Vol. 2049. Machine learning and its applications* (pp. 249–257). Springer. https://doi.org/10.1007/3-540-44673-7_12

Fernandino, G., Elliff, C. I., & Silva, I. R. (2018). Ecosystem-based management of coastal zones in face of climate change impacts: Challenges and inequalities. *Journal of Environmental Management*, *215*, 32–39. https://doi.org/10.1016/j.jenvman.2018.03.034

Forkuor, G., Dimobe, K., Serme, I., & Tondoh, J. E. (2018). Landsat-8 vs. Sentinel-2: Examining the added value of Sentinel-2's red-edge bands to land-use and land-cover mapping in Burkina Faso. *GIScience and Remote Sensing*, *55*(3), 331–354. https://doi.org/10.1080/15481603.2017.1370169

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, *29*(5), 1189–1232. https://doi.org/10.1214/aos/1013203451

Ghanem, S., Couturier, R., & Gregori, P. (2021). An accurate and easy to interpret binary classifier based on association rules using implication intensity and majority vote. *Mathematics*, *9*(12), Article 1315. https://doi.org/10.3390/math9121315

Goldblatt, R., Rivera Ballesteros, A., & Burney, J. (2017). High spatial resolution visual band imagery outperforms medium resolution spectral imagery for ecosystem assessment in the Semi-Arid Brazilian Sertão. *Remote Sensing*, *9*(12), Article 1336. https://doi.org/10.3390/rs9121336

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, *202*, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

Gupta, K. (2015). *Unprecedented growth of Dehradun urban area: A spatio-temporal analysis*. https://www.researchgate.net/publication/282334185_Unprecedented_growth_of_Dehradun_urban_area_a_spatio-temporal_analysis

Jin, S., Liu, X., Yang, J., Lv, J., Gu, Y., Yan, J., Yuan, R., & Shi, Y. (2022). Spatial-temporal changes of land use/cover change and habitat quality in Sanjiang plain from 1985 to 2017. *Frontiers in Environmental Science*, *10*, 1–12. https://doi.org/10.3389/fenvs.2022.1032584

Liu, C., Liu, Y., Lu, Y., Liao, Y., Nie, J., Yuan, X., & Chen, F. (2019). Use of a leaf chlorophyll content index to improve the prediction of above-ground biomass and productivity. *PeerJ*, *6*(1), Article e6240. https://doi.org/10.7717/peerj.6240

Liu, Y., Liu, L., & Yan, Y. (2020). Network topology change detection based on statistical process control. In *ACM International Conference Proceeding Series* (pp. 145–151). ACM. https://doi.org/10.1145/3409501.3409532

Nooni, I. K., Duker, A. A., Van Duren, I., Addae-Wireko, L., & Osei Jnr, E. M. (2014). Support vector machine to map oil palm in a heterogeneous environment. *International Journal of Remote Sensing*, *35*(13), 4778–4794. https://doi.org/10.1080/01431161.2014.930201

Osuna, E. E., Freund, R., & Girosi, F. (1999). *Support vector machines: Training and applications*. https://www.researchgate.net/publication/2592728_Support_Vector_Machines_Training_and_Applications

Pintelas, P., & Livieris, I. E. (2020). Special issue on ensemble learning and applications. *Algorithms*, *13*(6), Article 140. https://doi.org/10.3390/a13060140

Pontius, R. G., & Millones, M. (2011). Death to Kappa: Birth of quantity disagreement and allocation disagreement for accuracy assessment. *International Journal of Remote Sensing*, *32*(15), 4407–4429. https://doi.org/10.1080/01431161.2011.552923

Qiao, M., Wong, C., & Zheng, W. (2019). Sustainable urbanisation and community well-being in suburban neighbourhoods in Beijing, China. *International Journal of Community Well-Being*, *2*(1), 15–39. https://doi.org/10.1007/s42413-019-00019-9

Serwa, A. (2012). New method for feature reduction of MSS satellite bands to produce single equivalent band. *Al-Azhar University Engineering Journal*, *7*(1), 519–526.

Serwa, A., El-Nokrashy, M., O Ali, O., & Dief-Allah, M. A. M. (2010). New method to determine the optimum bands of MSS satellite images for unsupervised classification. *Al-Azhar University Engineering Journal*, *5*(1), 727–735.

Serwa, A., & Elbialy, S. (2021). Enhancement of classification accuracy of multi-spectral satellites' images using Laplacian pyramids. *Egyptian Journal of Remote Sensing and Space Science*, *24*(2), 283–291. https://doi.org/10.1016/j.ejrs.2020.12.006

Srivastava, A., & Ahmad, P. (2016). A probabilistic Gossip-based secure protocol for unstructured P2P networks. *Procedia Computer Science*, *78*, 595–602. https://doi.org/10.1016/j.procs.2016.02.122

Srivastava, A., & Biswas, S. (2023). Analyzing land cover changes over Landsat-7 data using Google Earth Engine. In *Proceedings of the 3rd International Conference on Artificial Intelligence and Smart Energy* (pp. 1228–1233), Coimbatore, India. https://doi.org/10.1109/ICAIS56108.2023.10073795

Srivastava, A., & Sharma, H. (2024). AI-driven environmental monitoring using Google Earth Engine. In B. Pradhan & S. Mukhopadhyay (Eds.), *Smart sensors, measurement and instrumentation: Vol. 50. IoT sensors, ML, AI and XAI: Empowering a smarter world* (pp. 375–385). Springer. https://doi.org/10.1007/978-3-031-68602-3_19

Srivastava, A., Bharadwaj, S., Dubey, R., Sharma, V. B., & Biswas, S. (2022). Mapping vegetation and measuring the performance of machine learning algorithm in Lulc classification in the large area using Sentinel-2 and Landsat-8 datasets of Dehradun as a test case. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLIII-B3-2022*, 529–535. https://doi.org/10.5194/isprs-archives-XLIII-B3-2022-529-2022

Srivastava, A., Dubey, R., & Biswas, S. (2023a). Comparison of Sentinel and Landsat data sets over Lucknow region using gradient tree boost supervised classifier. In A. Noor, K. Saroha, E. Pricop, A. Sen, & G. Trivedi (Eds.), *Lecture notes in networks and systems: Vol. 730. Proceedings of third emerging trends and technologies on intelligent systems* (pp. 221–232). Springer. https://doi.org/10.1007/978-981-99-3963-3_18

Srivastava, A., Umrao, S., & Biswas, S. (2023b). Exploring forest transformation by analyzing spatial-temporal attributes of vegetation using vegetation indices. *International Journal of Advanced Computer Science and Applications*, *14*(5), 1110–1117. https://doi.org/10.14569/IJACSA.2023.01405114

Srivastava, A., Umrao, S., Biswas, S., Dubey, R., & Zafar, M. I. (2023c). FCCC: Forest cover change calculator user interface for identifying fire incidents in forest region using satellite data. *International Journal of Advanced Computer Science and Applications*, *14*(7), 948–959. https://doi.org/10.14569/IJACSA.2023.01407103

Thenkabail, P. S., Teluguntla, P. G., Xiong, J., Oliphant, A., Congalton, R. G., Ozdogan, M., Gumma, M. K., Tilton, J. C., Giri, C., Milesi, C., Phalke, A., Massey, R., Yadav, K., Sankey, T., Zhong, Y., Aneece, I., & Foley, D. (2021). *Global cropland-extent product at 30-m resolution (GCEP30) derived from Landsat satellite time-series data for the year 2015 using multiple machine-learning algorithms on Google Earth Engine cloud* (Professional Paper No. 1868). US Geological Survey. https://doi.org/10.3133/pp1868

Viana, C. M., Girão, I., & Rocha, J. (2019). Long-term satellite image time-series for land use/land cover change detection using refined open source data in a rural region. *Remote Sensing*, *11*(9), Article 1104. https://doi.org/10.3390/rs11091104